

Fouille visuelle de données en 3D et réalité virtuelle : état de l'art

Zohra Ben Said*, Fabrice Guillet*, Paul Richard**

* LINA, UMR 6241 CNRS
(zohra.bensaid,fabrice.guillet)@univ-nantes.fr
www.lina.univ-nantes.fr

** LISA, EA 4094
paul.richard@istia.univ-angers.fr
www.istia.univ-angers.fr

Résumé. La fouille visuelle de données (ou Visual Data Mining, VDM) a pour objectif de faciliter l'interprétation des résultats issus d'une fouille de données, grâce à l'usage de représentations graphiques. Au cours de la dernière décennie, un grand nombre de techniques de visualisation d'information ont été mises au point, permettant la visualisation de données multidimensionnelles dans des environnements virtuels. Lors des travaux antérieurs, les chercheurs ont proposé des taxonomies pour classer les techniques de VDM (Chi (2000), Herman et al. (2000)). Toutefois, ces taxonomies ne prennent en compte que partiellement les techniques récentes relatives à l'utilisation de la 3D et de la réalité virtuelle. Le but de cet article est de faire un état de l'art récent et spécifique à ces techniques. Celles-ci sont détaillées, classées et comparées selon différents critères : les applications, l'encodage graphique, les techniques d'interaction, les avantages et les inconvénients de chaque approche. Ces techniques sont présentées dans des tableaux accompagnées d'illustrations graphiques.

1 Introduction

Une des limites sérieuses de la fouille de données provient des difficultés rencontrées par le décideur à interpréter les résultats produits. La fouille visuelle de données (visual Data Mining, VDM) propose de faciliter l'interprétation des résultats grâce à des représentations visuelles pertinentes. En effet, selon Card et al. (1999), l'esprit humain traite facilement les informations visuelles et en extrait rapidement un grand nombre de connaissances. Au début, les techniques de VDM étaient essentiellement basées sur des représentations 2D telles que les coordonnées parallèles (Inselberg et Dimsdale (1990)). Avec l'augmentation de la taille, de la complexité et de l'hétérogénéité des données stockées, de nouveaux paradigmes et métaphores de visualisation ont été développés. En effet, les résultats expérimentaux de Carswell et al. (1991) font état d'une amélioration des temps de réponse grâce à la 3D par rapport à la 2D pour la perception de tendances générales dans les données, nécessaires pour plusieurs applications telles que : (Tee et al. (2004), Schreck et al. (2007), Marroquín et al. (2008)). La 3D permet, grâce

à des scènes de plus en plus élaborées, de percevoir les données "de l'intérieur", ce qui augmente les possibilités de détecter les tendances, les clusters, etc (Keim et al. (2008)). Dans ce contexte, la navigation dans une scène 3D doit être intuitive et l'interaction doit se faire en temps réel. Dans ce cadre de visualisation interactive, l'évolution subjective des représentations graphiques par les utilisateurs montre également une préférence significative pour la 3D (Carswell et al. (1991)). Aussi, le travail collaboratif et le partage d'information (Simoff (2001)) sont plus faciles à réaliser en utilisant les techniques de réalité virtuelle (grand écran stéréoscopique).

Le but de cet article est de faire un état de l'art récent et spécifique à ces techniques. Celles-ci sont détaillées, classées et comparées selon différents critères : les applications, l'encodage graphique, les techniques d'interaction, les avantages et les inconvénients de chaque approche. Ces techniques sont présentées dans des tableaux avec des illustrations graphiques.

Cet article est organisé comme suit. La section 2 est dédiée à la technique focus + contexte. Dans la section 3 nous présenterons les systèmes de visualisation de mondes virtuels.

2 La technique focus/contexte

A l'origine, la méthode de visualisation focus + contexte (F + C), vise à fournir plus d'espace à la description détaillée de certains éléments de données (point d'intérêt, focus), tout en montrant le reste des données en taille réduite pour garder le contexte. Les techniques F + C les plus connues sont les techniques de distorsion (*fish-eye*) proposées par Furnas (1986).

Dans la technique *bending backwards*, une autre variante de la technique F + C, la vue d'ensemble de différents objets n'est pas lisible, mais leur aspect visuel miniature constitue un indice qui peut permettre de se déplacer directement vers l'information recherchée.

Cependant, il existe d'autres méthodes que la distorsion de l'espace. La visualisation de volume, par exemple, propose de faire varier l'opacité (Mroz et Hauser (2001)), les nuances de couleurs et la fréquence pour atteindre une visualisation F + C des données en 3D. Une comparaison détaillée des techniques F + C est présentée dans TAB.1.

3 Les mondes virtuels

Les *mondes virtuels* (parfois appelés *Cyber-espaces*) consistent à représenter l'information par des objets repartis dans un grand espace au sein duquel l'utilisateur peut naviguer (contrôler son point de vue). La visualisation des mondes virtuels en VDM repose sur la métaphore de *galaxie d'information* (Krohn (1996)) ou bien la métaphore de *paysage d'information* (Robertson et al. (1998)). Dans un paysage d'information, la dimension hauteur dans la position des objets n'est pas utilisée pour représenter une information. La technique de visualisation des nuages de points 3D repose sur la métaphore de *galaxie d'information*.

Les graphes dynamiques sont des métaphores de type *galaxie d'information* utilisés pour visualiser des graphes représentés par un système de forces qui agissent sur les noeuds et les liens.

Les arbres 3D permettent une visualisation d'information (données, concepts, etc.) organisée de manière hiérarchique. La connaissance au préalable de la forme de l'arbre par l'utilisateur l'aide à s'orienter dans l'espace de visualisation. Une comparaison des techniques de *mondes virtuels* est présentée dans TAB.2.

Systèmes de visualisation	Applications	Encodage graphique	Techniques d'interaction	Avantages	Inconvénients
Visualization Fish-eye -CbVAR Couturier et al. (2007) Fig.1(a)	Règles d'association	-2D : context -3D : focus	-Sélection -Zoom.	-L'affichage du contexte aide à l'orientation -Présentation des données sous forme de cluster -Paramétrage dynamique	-Peu de paramètres afficher
- Wang et al. (2008) Fig.1(b)	Forme 3D	-Agrandir le point de focus	-Sélection -Zoom	-Les données non focus gardent leurs tailles réels	-Contraintes dans le cas où il n'y a pas assez d'espace disponible
Bending backwards -3D-XV (Jacquemin et Jardino (2002)) Fig.1(c)	Structures linéaires	-La région focus dans le centre de l'écran et les données proches sur les côtés	-Navigation -Sélection	-Plusieurs modes d'accès à l'information -Orientation et déplacement direct facile	-Visualisation d'une sous-partie des données à la fois
Linking and brushing Couleur : WEAVE (Gresh et al. (2000)) Fig.1(d), SimVis (Doleisch et al. (2005)) Fig.1(e) Opacité : The Magic Volume Lens (Wang et al. (2005)) Fig.1(g), (Götzelmann et al. (2007)) Fig.1(h) Fréquence : (Elmqvist et al. (2009)) Fig.1(i)	Données médicales, scientifiques et industrielles	-Utilisation irrégulière des ressources graphiques	-Sélection -Feedback (changement de couleur, cli-gnotement...) -Des vues multiples liées	-Changements interactifs immédiats -Détection rapide des dépendances et des corrélations	-Pas de zoom sémantique.

TAB. 1 – Tableau comparatif des techniques Focus + contexte

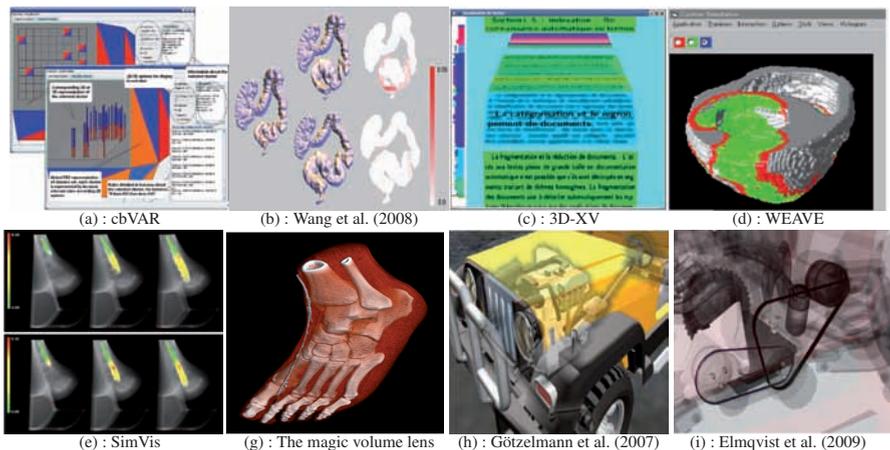
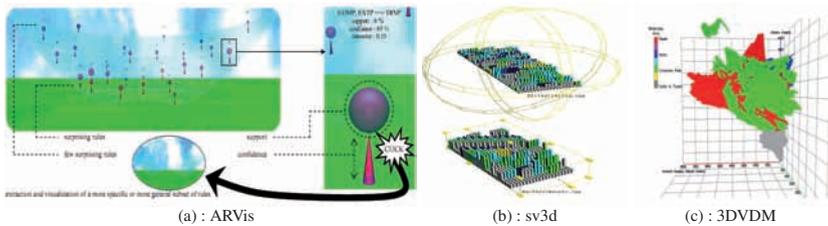


FIG. 1 – Illustrations des techniques Focus + contexte

Fouille visuelle de données en 3D et réalité virtuelle

Systèmes de visualisation	Applications	Encodage graphique	Techniques d'interaction	Avantages	Inconvénients
Les mondes virtuels					
-ARVis Blanchard et al. (2007) Fig.2(a)	-Règles d'association	-Taille du cône : confiance -Taille de sphère : support -Position : intensité d'implication -Couleur : moyenne des 3 mesures	-Navigation -Sélection -Zoom -Filtre	-Visualisation par ensemble de règles -Navigation selon les relations de voisinage	-Pas de représentation hiérarchique -Pas de vue globale du contexte
-Source Viewer 3D (sv3D)Maletic et al. (2003) Fig.2(b)	-Structure de code source	-Conteneur : fichier -Couleur : type de la structure de contrôle	-Navigation -Sélection -Zoom -Filtre	-Historique -Les imprimés-écrans -Libre	-Pas de relations entre les classes ou les fichiers -Pas de représentation hiérarchique
Nuages de points 3D					
-3D Scatter Plot(VR)(Bovbjerg et al. (2003)) Fig.2(c)	-Grande base de données	-Couleurs/textures : objets/clusters	-Navigation -Zoom -Sélection	-Utilisation du son	-Nombre limité de variables graphiques -Pas très performant
-VRMiner(VR) (Azzag et al. (2005)) Fig.2(d)	-Données multimédias	-Position, forme, taille, couleur, son et texture	-Navigation -Zoom -Sélection -Synthèse vocale	-Visualisation des images de grande taille -Faible coût -Libre	-Nombre limité de variables graphiques
Les graphes auto-organisés					
-PEX (Paulovich et al. (2007)) Fig.2(e)	-Données multidimensionnelles	-Couleur : fréquence d'apparition	-Recherche -Sélection - Personnalisation -Filtrage	-Libre -Visualisation des données structurées et non structurées	-Pas de détails sur demande
Les arbres 3D					
-OntoSphere3D (Bosca et al. (2007)) Fig.2(f) -SUMO (Buntain (2008))	-Visualisation des ontologies	-Sphère : concept -Cube : instance -Lien : relation sémantique	-Rotation -Navigation -Zoom -Détails sur demande	-Facilement interprétable	-Interaction avec une seule main

TAB. 2 – Tableau comparatif des techniques de mondes virtuels



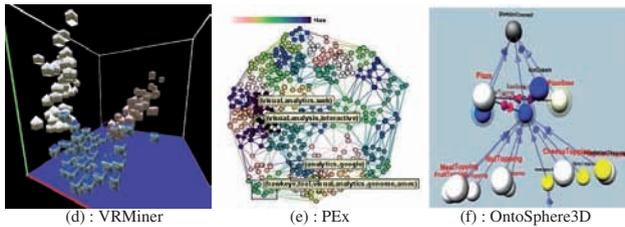


FIG. 2 – Illustrations des techniques de mondes virtuels

4 Conclusion

Si les travaux sur la fouille de données sont relativement anciens, le domaine de la fouille visuelle de données (VDM) l'est moins. La VDM cherche à exploiter au mieux les capacités sensori-motrices et cognitives de l'être humain à l'aide de techniques interactives afin que l'utilisateur puisse extraire des connaissances que la machine n'aurait pu découvrir. L'évaluation des techniques 3D reste encore un problème ouvert.

Références

- Azzag, H., F. Picarougne, C. Guinot, et G. Venturini (2005). Vrminer : a tool for multimedia databases mining with virtual reality. *Processing and Managing Complex Data for Decision Support*, 318–339.
- Blanchard, J., B. Pinaud, P. Kuntz, et F. Guillet (2007). Visual analytics : A 2d-3d visualization support for human-centered rule mining. *Computers and Graphics* 31(3), 350–360.
- Bosca, A., D. Bonino, M. Comerio, S. Grega, et F. Corno (2007). A reusable 3d visualization component for the semantic web. In *Web3D '07 : Proceedings of the twelfth international conference on 3D web technology*, pp. 89–96. ACM Press.
- Bovbjerg, S., E. Granum, H. R. Nagel, et M. Vittrup (2003). Using dynamic soundscapes to support visual data mining in vr. In M. H. B. Simeon J. Šimoff, Monique Noirhomme-Fraiture et M. I. Ankerst (Eds.), *Third International Workshop on Visual Data Mining in conjunction with ICDM 2003 - The Third IEEE International Conference on Data Mining*, pp. 167–182.
- Buntain, C. (2008). 3d ontology visualization in semantic search. In *Proceedings of the 46th Annual Southeast Regional Conference on ACM Southeast Regional Conference*, pp. 204–208. ACM Press.
- Card, S. K., J. D. Mackinlay, et B. Schneiderman (1999). *Readings in information visualization : using vision to think*. Morgan Kaufmann publishers, San Francisco CA, ETATS-UNIS (Monographie).
- Carswell, C. M., S. Frankenberger, et D. Bernhard (1991). Graphing in depth : Perspectives on the use of three-dimensional graphs to represent lower-dimensional data. *Behaviour & Information Technology*. 10(6), 459–474.
- Chi, E. H. (2000). A taxonomy of visualization techniques using the data state reference model. In *INFOVIS '00 : Proceedings of the IEEE Symposium on Information Visualization*, pp. 69–75. IEEE Computer Society Press.
- Couturier, O., T. Hamrouni, S. B. Yahia, et E. M. Nguifo (2007). A scalable association rule visualization towards displaying large amounts of knowledge. In *IV '07 : Proceedings of the 11th International Conference Information Visualization*, pp. 657–663. IEEE Computer Society Press.
- Doleisch, H., M. Mayer, M. Gasser, P. Priesching, et H. Hauser (2005). Interactive feature specification for simulation data on time-varying grids. In *Conference on Simulation and Visualization*, pp. 291–304. SCS Publishing House e.V.
- Elmqvist, N., U. Assarsson, et P. Tsigas (2009). Dynamic transparency for 3d visualization : Design and evaluation. *The International Journal of Virtual Reality* 8(1), 75–88.
- Furnas, G. W. (1986). Generalized fisheye views. In *CHI '86 : Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 16–23. ACM Press.
- Gresh, D. L., B. E. Rogowitz, R. L. Winslow, D. F. Scollan, et C. K. Yung (2000). Weave : a system for visually linking 3-d and statistical visualizations, applied to cardiac simulation and measurement data. In *Proceedings of the conference on Visualization '00*, pp. 489–492. IEEE Computer Society Press.

- Götzelmann, T., K. Hartmann, A. Nürnberger, et T. Strothotte (2007). 3d spatial data mining on document sets for the discovery of failure causes in complex technical devices. In *GRAPP'07 : Proceedings of the Second International Conference on Computer Graphics Theory and Applications*, pp. 137–145. INSTICC - Institute for Systems and Technologies of Information, Control and Communication.
- Herman, I., G. Melançon, et M. S. Marshall (2000). Graph visualization and navigation in information visualization : A survey. *IEEE Transactions on Visualization and Computer Graphics* 6(1), 24–43.
- Inselberg, A. et B. Dimsdale (1990). Parallel coordinates : a tool for visualizing multi-dimensional geometry. In *Proceedings of the 1st conference on Visualization*, pp. 361–378. IEEE Computer Society Press.
- Jacquemin, C. et M. Jardino (2002). Une interface 3d multi-échelle pour la visualisation et la navigation dans de grands documents xml. In *IHM '02 : Proceedings of the 14th French-speaking conference on Human-computer interaction (Conférence Francophone sur l'Interaction Homme-Machine)*, pp. 263–266. ACM Press.
- Keim, D. A., F. Mansmann, J. Schneidewind, J. Thomas, et H. Ziegler (2008). *Visual Analytics : Scope and Challenges*. Berlin, Heidelberg : Springer-Verlag.
- Krohn, U. (1996). Vineta : navigation through virtual information spaces. In *AVI'96 : Proceedings of the workshop on Advanced visual interfaces*, pp. 49–58. ACM Press.
- Maletic, J. I., A. Marcus, et L. Feng (2003). Source viewer 3d (sv3d) : a framework for software visualization. In *ICSE'03 : Proceedings of 25th ACM/IEEE International Conference on Software Engineering*, pp. 812–813. IEEE Computer Society Press.
- Marroquin, V. D., J. J. Brault, et B. S. Hart (2008). A visual data-mining methodology for seismic facies analysis : Part 2 - application to 3d seismic data. *GEOPHYSICS* 74(1), 13–23.
- Mroz, L. et H. Hauser (2001). Rtvr : a flexible java library for interactive volume rendering. In *VISUALIZATION'01 : Proceedings of the Conference on Visualization*, pp. 279–286. IEEE Computer Society Press.
- Paulovich, F. V., M. C. F. Oliveira, et R. Minghim (2007). The projection explorer : A flexible tool for projection-based multidimensional visualization. In *SIBGRAPI'07 : Proceedings of the Brazilian Symposium on Computer Graphics and Image Processing*, pp. 27–36. IEEE Computer Society Press.
- Robertson, G., M. Czerwinski, K. Larson, D. C. Robbins, D. Thiel, et M. van Dantzich (1998). Data mountain : using spatial memory for document management. In *Proceedings of the 11th annual ACM symposium on User interface software and technology*, pp. 153–162. ACM Press.
- Schreck, T., T. Tekušová, J. Kohlhammer, et D. Fellner (2007). Trajectory-based visual analysis of large financial time series data. *ACM SIGKDD Explorations Newsletter* 9(2), 30–37.
- Simoff, S. (2001). Towards the development of environments for designing visualisation support for visual data mining. In *Proceedings International Workshop on Visual Data Mining*, pp. 93–106. Simeon J. Simoff, Monique Noirhomme-Fraiture, Michael H. Böhlen and Mihael I. Ankerst.
- Tee, S., T. T. J. J. Kelly Kwan-liu Ma, et S. F. Wu (2004). Visual data analysis for detecting flaws and intruders in computer network systems. *IEEE Computer Graphics and Applications, special issue on Visual Analytics* 24(5), 27–25.
- Wang, L., Y. Zhao, K. Mueller, et A. Kaufman (2005). The magic volume lens : An interactive focus+context technique for volume rendering. In *VIS 05 : Proceeding of 16th IEEE Visualization*, pp. 367–374. IEEE Computer Society Press.
- Wang, Y.-S., T.-Y. Lee, et C.-L. Tai (2008). Focus+context visualization with distortion minimization. *IEEE Transactions on Visualization and Computer Graphics* 14(6), 1731–1738.

Summary

The visual data Mining (VDM) aims at improving the interpretation of data mining results through the use of visualization techniques. During the last decade, many techniques of information visualization have been developed, enabling visualization of multidimensional data in virtual environments. In previous work, researchers have proposed taxonomies for classifying VDM techniques (Chi (2000), Herman et al. (2000)). However, these taxonomies do not take completely into account the latest techniques such as 3D and virtual reality. The purpose of this paper is to review new and specific survey to these techniques. These are detailed, classified and compared according to different criteria: applications, graphical encoding, interaction techniques, advantages and disadvantages of each approach. These techniques are presented in tables accompanied by graphic illustrations.